Research Report

Design Strategies and Performance of Custom DNA Sequencing Primers

BioTechniques 27:528-536 (September 1999)

G.A. Buck[1], J.W. Fox[2], M. Gunthorpe[3], K.M. Hager[4], C.W. Naeve[5], R.T. Pon[6], P.S. Adams[7] and J. Rush[8]

[1]Virginia Commonwealth University, Richmond, VA; [2]University of Virginia, Charlottesville, VA; [3]University of California, San Francisco, CA; [4]Yale University, New Haven, CT; [5]St. Jude Children's Research Hospital, Memphis, TN, USA; [6]University of Calgary, Calgary, AB, Canada; [7]Trudeau Institute, Saranac Lake, NY and [8]Harvard Medical School, Boston, MA, USA

ABSTRACT

This study surveyed strategies of sequencing primer selection and evaluated primer performance in automated DNA sequencing. We asked participants to relate their preferred primer design strategies to identify primer characteristics that are considered most important in sequencing primer design. The participants preferred primers of 18–24 nucleotides (nt), 39%–58% G+C, a melting temperature (T_m) of 53°–65°C with a 1–2 nt 3′ GC clamp, hairpin stems of less than 2–3 bp, homopolymeric runs of less than 4–5 nt, primer dimers of less than 3–4 bp and secondary priming sites of less than 3–4 bp. We provided a 300-bp test sequence and asked participants to submit sequences of 1–3 optimal sequencing primers. Submitted primers ranged from 17–24 nt and largely conformed to the preferred parameters. Submitted primers were distributed across the test sequence, although some sites were disfavored. Surprisingly, approximately 45% of the primers were selected "manually", more than by any software package. Each of 69 submitted and 95 control primers, distributed at 3 bp intervals across the test sequence, were synthesized, purified and tested using a Model 377 PRISM™ DNA Sequencer with dichlororhodamine dye terminator reagents (dRhodamine dye terminators). Approximately half of the control primers were also tested using rhodamine dye terminator reagents ("old" rhodamine dye terminators). The results indicated that primer physico-chemical characteristics thought to have a strong impact on sequencing performance had surprisingly little effect. Thus, primers with high or low percent G+C or T_m, strong secondary priming scores or long 3′ homopolymeric stretches yielded excellent sequences with the dRhodamine dye terminator reagents, although these characteristics had a stronger effect when the old rhodamine reagents were used. The old rhodamine reagents gave sequences with a similar average read length, but the number of errors and ambiguities or "N's" was consistently higher. Moreover, the effects of the primer physico-chemical characteristics were also more evident with the old rhodamine dyes. We conclude that under optimal sequencing conditions with highly pure template and primer, many of the commonly applied primer design parameters are dispensable, particularly when using one of the new generation of sequencing reagents such as the dichlororhodamine dye terminators.

INTRODUCTION

The Nucleic Acids Research Committee (NARC) of the Association of Biomolecular Resource Facilities (ABRF) evaluates the procedures and performance of research facilities that perform nucleic acid synthesis and sequencing in academic, industrial and other institutions (5,6). These studies provide a measure of the industry state-of-the-art, as well as the level of performance, that can be expected from such facilities. Previous projects have examined general operations of DNA core laboratories (15), the accuracy of automated DNA sequencing in these facilities (12), and the performance of unpurified synthetic oligonucleotides as primers for automated DNA sequence analysis (14). The DNA Sequencing Research Committee of the ABRF has performed studies evaluating the ability of DNA core facilities to sequence difficult templates (2) and the performance of various sequencing strategies and chemistries (1). Herein, we have studied primer selection strategies used by DNA sequencing facilities and empirically examined the quality of the sequence data provided by primers selected by these facilities.

Despite a long-standing conviction within the DNA sequencing community that careful primer design is essential to ensure high quality data (c.f., 11,14,17,18), there is a paucity of empirical studies supporting this view. A variety of reasonable "rules of thumb" (13) suggesting optimal ranges for primer length, percent G+C, melting

Table 1. Predicted Optimal and Observed Physico-Chemical Characteristics of Primers for Automated DNA Sequencing or PCR

	Sequencing Primers[a]		PCR Primers[b]		Submitted Primers[c]	
	Optimum[d]	Range[e]	Optimum[d]	Range[e]	Mean[f]	Range[g]
Length	18–24 nt	15–40 nt	18–29 nt	15–40 nt	19.7 nt	16–24 nt
T_m[h]	52°–66°C	40°–95°C	50°–69°C	40°–95°C	66°C	48°–91°C
Percent G+C of primer	40%–61%	30%–70%	40%–61%	30%–70%	53%	29%–82%
3′ GC clamp[i]	1–2 nt	0–5 nt	1–2 nt	0–2 nt	1.1 nt	0–4 nt
Primer dimer[j]	3–4 bp	0–7 bp	3–4 bp	0–7 bp	-5.4	-1.6–16.5
Hairpin stems[k]	2–3 bp	0–6 bp	2–3 bp	0–6 bp	1.1 bp	0–5 bp
Homopolymer runs[l]	4–5 nt	3–10 nt	4–5 nt	3–10 nt	3.0 nt	2–5 nt
Secondary priming[m]	3–4	0–10 bp	3–4 bp	0–10 bp	49[n]	0–266[n]

[a]Parameter values preferred by participants for primers to be used in sequencing.
[b]Parameter values preferred for primers to be used in PCR experiments.
[c]Parameter values of the primers submitted by participants.
[d]Optimum values for each parameter as selected by the participants.
[e]The range of acceptable values selected by the participants.
[f]The mean values for each parameter calculated from the submitted primers.
[g]The range of values for each parameter calculated from the submitted primers.
[h]Values submitted by participants for primers to be used for sequencing or PCR and nearest neighbor T_m of primers submitted by participants.
[i]The 3′ clamp refers to the number of G or C bases at the 3′ terminus of the primer.
[j]Largest number of contiguous complementary bases permitting dimerization of primer or the stability of strongest potential primer dimer (kcal) as measured by Oligo 5.0.
[k]Largest number of bases capable of forming a hairpin stem in the primer.
[l]Number of bases in longest homolopolymeric run in the primer.
[m]Number of bases permitted in most stable secondary priming site.
[n]The priming efficiency as calculated by Oligo 5.0 for test sequence only (submitted primers).

temperature (T_m) and composition have led to the development of primer design software packages that identify primers conforming to these criteria (4,16,18). However, the success of these strategies in selection of high-quality sequencing primers, to our knowledge, has not been extensively examined. In many sequencing projects, primer design and synthesis represent the most significant costs and consume the bulk of effort and time. Therefore, it is extremely important that these primers yield high-quality sequence data.

To examine this issue, an e-mail survey containing general questions about laboratory functions and specific questions concerning important characteristics for primer selection was distributed. A 300-bp test sequence was provided for which participants were asked to design sequencing primers. The submitted primers and a set of control primers spanning the 300-bp test sequence at 3-bp intervals were synthesized and used to sequence the test template on a Model 377 PRISM™ Automated DNA Sequencer (PE Biosystems, Foster City, CA, USA). Due to the results of this study, current primer design rules of thumb may be streamlined to facilitate more efficient primer selection for sequencing projects.

MATERIALS AND METHODS

The Survey

In December of 1996 and January of 1997, members of the ABRF were requested via e-mail distributions, the ABRF Electronic Bulletin Board and a standard mailing, to participate in a study of DNA sequencing primer design. The distribution provided a 300-bp "test sequence" (Figure 1) that was selected and previously shown to contain no segments that affect sequence ladder extension (data not shown). Participants were asked to use the prevailing technology in their laboratories to design and submit the sequences of up to three sequencing primers in the forward direction. Participants were requested to select the best primers irrespective of their positions on the test sequence. All responses were anonymously screened and coded by Virginia Commonwealth University (VCU) Health Sciences Computer Center personnel. Identifying data were removed from each document before being forwarded to the Committee for analysis. Several respondees submitted uninterpretable data. Where possible, these respondees were identi-

fied by Computer Center personnel, and new information was requested. Every effort was made to maintain the integrity of the data and the anonymity of the participants.

Control and Submitted Primers

A panel of 95 primers were synthesized as controls. These 18 nucleotide (nt) primers spanned the 300-bp test sequence with their 5′ termini located at 3-bp intervals (i.e., the first began at bp 1 of the test sequence, the second began at bp 4, etc.). All primers were analyzed with the Oligo™ Program 5.0 (NBI/Genovus, Plymouth, MN, USA) (16) and the number of bases, percent G+C, T_m (nearest neighbor method described in Reference 9), internal structure, secondary priming sites, etc., were recorded.

The 69 submitted and 95 control primers were synthesized on a Model 3948 DNA Synthesis and Purification System (PE Biosystems) at three different sites. Synthesis was performed at the 40 nmol scale using PE Biosystems reagents and the standard synthesis, cleavage and purification cycles. An aliquot of each sample was analyzed by polyacrylamide gel electrophoresis (PAGE) and by matrix-assisted laser desorption ionization time of flight mass spectrometry (MALDI-TOF/MS). PAGE was performed using precast minigels (Novex, San Diego, CA, USA) following the protocol specified by the manufacturer. Mass analysis was performed on a 200-pmol sample mixed with 10 mg/mL 3-hydroxypicolinic acid matrix (in 50% acetonitrile, 0.1% trifluoracetic acid) on a Voyager RP MALDI-TOF/MS instrument (PerSeptive Biosystems, Framingham, MA, USA). Criteria for passing quality control were: (*i*) a single major band on the gel; (*ii*) appropriate mobility on the gel compared to molecular weight standards; (*iii*) a synthesis yield >0.5 ODU (absorption of sample in 1 mL at 260 nm with 1 cm pathlength set up on the 3948 Synthesizer); and (*iv*) a single peak of the expected mass by MALDI-TOF/MS. Any oligonucleotides that failed to meet these criteria were resynthesized.

Test Template and Preparation

The plasmid template was preselected to contain a test sequence lacking obstacles to sequence extension (data not shown) and purified by double banding in CsCl-ethidium bromide isopycnic density gradients (10).

DNA Sequence Analysis

Each of the purified and quality-tested oligonucleotides were used as a primer for DNA sequence analysis using dRhodamine Terminator Cycle Sequencing Ready Reaction Kit with AmpliTaq® FS reagents (PE Biosystems) under standard reaction conditions as described by the manufacturer at 20 µL total vol containing 300 ng of template and 5 pmol of primer. Sequencing reactions were run for 25 cycles of 10 s at 96°C, 5 s at 50°C and 4 min at 60°C in a GeneAmp® PCR System 9700 DNA Thermal Cycler (PE Biosystems). Reactions were analyzed on Model 377 PRISM Automated DNA Sequencers using 5% Long Ranger™ (FMC BioProducts, Rockland, ME, USA) gels with Tris-Borate-EDTA buffer (TBE; pH 8.3). Approximately 30% of the control primers were rerun using the same dRhodamine dye terminator/Taq FS kit to ensure reproducibility, and approximately 50% of the control primers were rerun using the old rhodamine dye terminator/Taq FS kit (PE Biosystems). The dRhodamine or dichlororhodamine reagents are modifications of the original old rhodamine dyes that have narrower emission spectra, less spectral overlap and more even peak heights than the earlier dyes (19). Each 5% polyacrylamide sequencing gel was electrophoresed for approximately 8 h, and the data was directly exported for analysis using Sequencher™ (Gene Codes, Ann Arbor, MI, USA). Using this software, the raw data was trimmed to remove 5′ and 3′ ambiguous sequences, so that the first and last 25 nt contained no ambiguities. Each result was aligned with the known sequence, and we recorded: (*i*) the number of nt between the 3′ end of the primer and the beginning of the trimmed data; (*ii*) the number of nt in each sequence read; and (*iii*) the number of errors, including miscalls, insertions, deletions and ambiguities in each 100-bp window through the end of the sequence read.

```
  1   ATCCACACGG CGCTTTGACT CCCCTTCTGA AAAAAAAGAA CACGTTGATT
 51   CCCCTTTTGA TTTCCCACGA ATGCGAAGCA CAATTCACGC ACAGCCAACG
101   GCAGGGAGAG ACGATAAGGT AAAATACACA CAAGTGGTGG AATCACAAAA
151   GGACACGCAC AGCACGCAAA TAAACATTGC ACACGGCATG ACATTTTGCG
201   ACCCACACAC AATGTGTACA CTACAATTAT ATGAAAATCC CTCCCCCATT
251   CCGCGGGTGC GCCGCAAAGG GCCAAAAACC CAAATGATCC ACTTTATTAT
```

Figure 1. Test sequence. Shown is the 300-bp sequence distributed to participants for design of forward sequencing primers. Participants were asked to design primers in the forward direction.

RESULTS

Participant Profiles

Thirty-nine laboratories participated in the study, submitting 69 primer sequences. Of the participants, approximately 77% provide synthesis services and approximately 85% provide sequencing services, but only approximately 50% provide primer design or walking services, and less than one third offer template preparation. Laboratories performing these services averaged annually approximately 3700 syntheses, 7000 sequence runs, 425 primer designs, 424 template purifications and less than 200 primer walking projects. Charges for services averaged $1.17/nt for 40–50 nmol and $2.19/nt for 150–200 nmol syntheses of a 25-nt unpurified primer and $24/reaction for DNA sequence analysis ($9 for gel only). The average price for a primer walking sequencing project of a 2-kbp insert in a plasmid vector was approximately $500 or $0.25/bp. Charges for "external users", where available, averaged approximately 17%–50% higher than charges for institutional users. DNA synthesis and sequencing instrumentation was dominated by PE Biosystems; i.e., 77% of synthesizers and 92% of sequencers, but Beckman

Instruments, PerSeptive Biosystems and LI-COR were also represented. Only approximately 15% of the participants maintained robotic systems for template preparation or sequencing. Finally, when asked if commercial suppliers had an impact on services provided by the facilities, almost 80% of the laboratories responded affirmatively. Approximately 38% had dropped

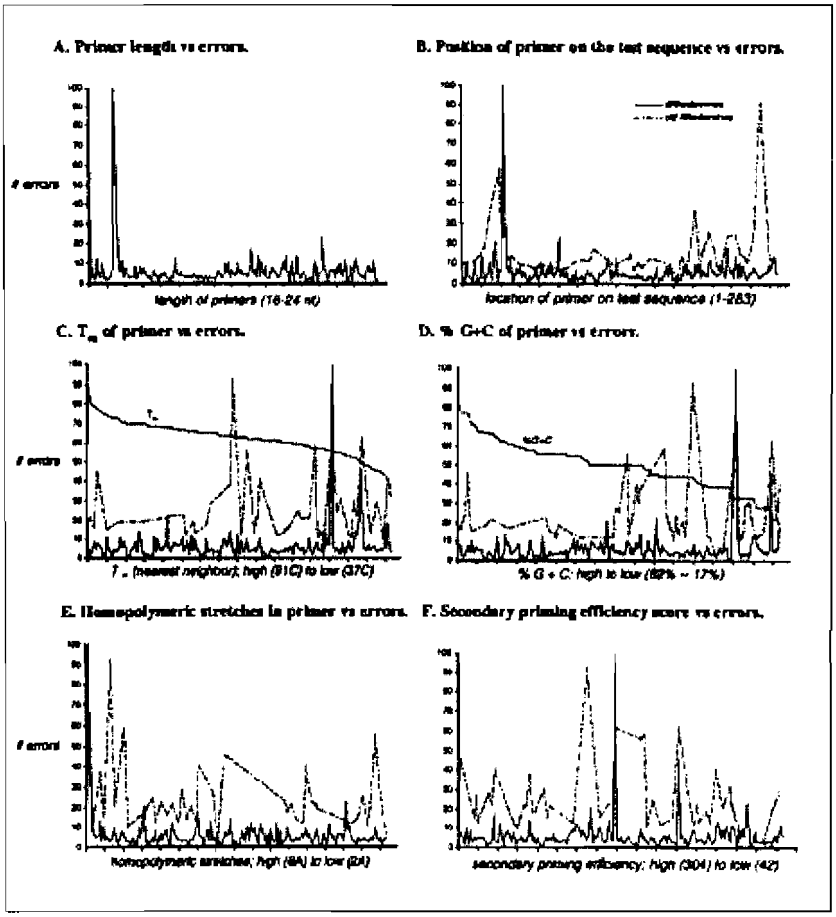

Figure 2. Primer characteristics vs. sequencing performance. The performance of sequencing primers, sorted according to various physico-chemical characteristics, was examined. The number of errors over the full-length read of each primer, determined as described in the Materials and Methods, was plotted against the results of each primer sorted according to the relevant parameter. (A) Primer length vs. errors. The primers were sorted according to their lengths in ascending order, and secondarily, according to their positions on the test sequence. The number of errors over the full-length read is plotted. Note that since all of the primers tested with the old rhodamine reagents were controls, and therefore 18-nt long, the data for old rhodamines are not included in this analysis. The X-axis distributes the primers from 16–24 nt in length; the y-axis indicates the number of errors over the full-length read of that primer. (B) Position of the primer on the test sequence. The primers were sorted according to the position of their 5′ nt on the test sequence (bp 1–283 on x-axis) and the number of errors in their full-length reads were plotted (y-axis). (C) T_m of the primer. The primers were sorted in descending order by their T_m's (91°–37°C on x-axis) and the number of errors in their full-length reads were plotted (y-axis). (D) The percent G+C of the primer. The primers were sorted in deseending order (82%–17% on x-axis) by their percent G+C, and the number of errors in their full-length reads were plotted (y-axis). (E) Homopolymeric stretches in the primer. The primers were sorted according to first the number of nt (8–2 on x-axis) in the longest homopolymeric streteh, and secondarily aceording to the number of nt in the second longest homopolymeric stretch. The number of errors over the full-length read were plotted (y-axis). (F) Secondary priming efficiency of the primers. Each primer was examined for secondary priming sites in the veetor and across the test sequence using Oligo 5.0. The priming efficieney score assigned by Oligo 5.0 was determined. The primers were sorted first according to the score of the strongest secondary priming site in the vector (304–42 on x-axis) and secondarily according to the strongest secondary priming score in the test sequence. The number of errors across the full-length reads were plotted (y-axis).

Table 2. Average Primer Performances of Submitted and Control Primers Using dRhodamine or Old Rhodamine Dye Terminator Sequencing Kits[a]

	Submitted Primers dRhodamines[b]	Control Primers dRhodamines[b]	Control Repetition dRhodamines[b]	Control Primers old kit dyes[b]
5′ nt trimmed[c]	3.3 (2, 1–14)	4.6 (4, 1–19)	4 (3, 0–16)	4.5 (4, 0–16)
read length in nt[d]	760 (770, 674–802)	772 (778, 683–799)	798 (803, 733–867)	752 (764, 654–803)
ambiguities (N's)[e]	2.2 (1, 0–16)	1.6 (1, 0–32)	4.4 (4, 1–17)	8.7 (7, 0–58)
errors per first 700 nt[f]	2.9 (2, 0–14)	2.9 (2, 0–46)	5.4 (5, 3–16)	16.5 (12, 3–92)

[a]Old rhodamine kits refer to the original ABI PRISM Dye Terminator Cycle Sequencing Kits, dRhodamine kits refer to the new ABI PRISM dRhodamine Terminator Cycle Sequencing Ready Reaction Kits containing the dichlororhodamine dye terminators.
[b]The data is shown as the average with the (median, range) in parentheses.
[c]The number of nt between the 3′ end of the primer and the beginning of the trimmed sequence.
[d]The length of read in nt after trimming using Sequencher (see Materials and Methods).
[e]The number of ambiguous calls (N's) in the full-length sequence read.
[f]The number of errors in the first 700 nt of the sequence.

prices and approximately 25% indicated a decreased number of synthesis requests in response to commercial suppliers. About 23% of the laboratories purchased or planned to purchase large-scale synthesis instruments to match the new competition. However, only approximately 5% had discontinued synthesis services.

Software or Manual Selection of Primers

Thirty-one laboratories described their practices for the design of sequencing and PCR primers. We were surprised that, in spite of the availability of excellent commercial primer design software programs, over 50% (17/31) of the reporting laboratories preferred to select primers manually. Of the 25 laboratories reported that selection and design of PCR primers, again over 50% (14/25) tended not to use software. Of those laboratories that rely on software for primer selection, over 60% rely on various versions of the Oligo software for sequence primer design, and over 70% use the Oligo software for PCR primer design.

Idealized Primer Physico-Chemical Characteristics

Participants were asked to provide their preferences for the physico-chemical characteristics of the ideal sequencing and PCR primers. This analysis provided few surprises (Table 1), in that the idealized primer characteristics of most respondents closely parallel those of some common rules of thumb for primer design. Thus, sequencing primers should be: 18–24 nt in length, T_m = 52°–66°C, 40%–60% G+C, with 1–2 bp 3′ GC clamps (G or C nt at 3′ terminus of the primer) and very little internal structure, homopolymeric sequence or potential for secondary priming. A similar set of characteristics was reported for PCR primers, although the optimal acceptable length, T_m and percent G+C were slightly greater than for sequencing primers. However, it was unexpected that some participants accept primers of up to 40 bases, with T_m's as high as 95°C for both sequencing and PCR primers. When asked to rate the relative importance of each characteristic, participants rated the existence of secondary priming sites and percent G+C as most important, with T_m, presence of a 3′ GC clamp, the potential to form primer dimers, primer length, the presence of internal structure or the presence of homopolymeric runs, in order of decreasing importance for sequencing primer design. Relative importance of these characteristics was similar for PCR primers, except that the potential for formation of primer dimers was rated almost as important as the existence of potential secondary priming sites.

Characteristics of the Submitted Primer Sequences

The 39 participating laboratories submitted 69 primer sequences ranging in length from 16–24 nt with an average of approximately 20 nt. The primers were distributed fairly evenly across the 300-bp test sequence, although some segments were obviously purposefully excluded; e.g., any primer beginning between bp 21 and 30 of the test sequence would bear an 8-nt poly(A) segment (not shown). Not surprisingly, most of the submitted primers conformed closely to the idealized physico-chemical characteristics for sequencing primers (Table 1). Although some deviated from the idealized characteristics, all conformed to the more broad ranges defined by participants. Thus, the participants selected primers according to their idealized criteria, even though almost half of the submitted primers were selected manually without the aid of computer software. Interestingly, some of the submitted primers had very high or low T_m (91° or 48°C) or percent G+C (82% or 29%), indicating that the participants do in fact permit a significant variation in these characteristics. Moreover, some primers exhibited relatively high

secondary priming efficiency numbers (a score calculated by Oligo 5.0, whereby a priming efficiency of 160 can prime in sequencing reactions) even within the test sequence (Table 1).

Sequencing Performance of the Primers

The results of the sequence analysis as described in the Materials and Methods were unexpected (Table 2); i.e., we expected a wide range in the performance of the submitted and control primers. Instead, all of the submitted primers functioned extremely well, the poorest performer still yielding <2% errors over a >700-nt sequence. The average number of errors and ambiguities over the 700-nt window was 2.9, and the average read length was 760 nt. Similarly, almost all of the control primers functioned very well, with an average of 2.9 errors or ambiguities over the 700-nt window and an average read length of 772 bases. Only one primer (No. 8) failed to give any reasonable sequence, and even this oligonucleotide performed adequately on the repeat; i.e., 13 errors and ambiguities over a 700-nt read. The average read lengths of the sequences from the submitted and control primers were 760 and 772, with ranges of 674–802 and 683–799 bases, respectively. The average number of nucleotides trimmed from the 5′ ends of these sequences was 3.3 and 4.6 nt, respectively. Many sequences began on the first nucleotide after the primers, but up to 14 or 19 nt were trimmed from the 5′ ends of the submitted and control primers. As expected from previous studies (1,2,12, 14,18), errors and ambiguities were clustered in the first and last 100-bp windows of the sequences (data not shown). Approximately 30% of the control primers were used in repeat reactions with very similar results (Table 2, and data not shown).

The control primers used in sequence reactions using the old rhodamine dye terminator reagents gave slightly poorer overall results. The read lengths of these reactions did not differ greatly from the read lengths of the dRhodamine reactions (average 752 nt vs. 772 bases, Table 2). However, there were more N's (average 8.7 vs. 1.6) and errors and ambiguities (average 16.5 vs. 2.9) in the data generated with old rhodamine dye terminators than in the dRhodamine-generated data (Table 2).

Poor Primers Yield Good Sequence Results

The uniform high-quality sequence data from the control primers using the dRhodamine reagents was not expected. Only primer No. 8 failed catastrophically (i.e., gave no valid sequence data), and primer No. 9 yielded >5% errors over a 700-nt window. However, both of these primers provided better data when used in a repeat sequencing reaction (Table 3). Moreover, there was no obvious relationship between the number of errors in a sequence and the length, position, T_m, percent G+C, secondary priming potential, length of homopolymeric stretches (Figure 2, panels A–F) or the potential for forming primer dimers (data not shown) of the primers. On the contrary, many primers exhibiting suboptimal physico-chemical characteristics yielded high-quality sequence. For example, although the one primer, No. 8, that failed castastrophically has an 8-nt poly(dA) stretch near its 3′ terminus, other primers, e.g., Nos. 7 and 10, contain a similar sequence, but still generated very respectable data (Table 3). Primer No. 9, which also contains the same poly(dA) stretch, also yielded relatively poor data (46 errors in 700 bases), although the error rate was <3% over the first 500-nt window (data not shown). Examination of the sequences of these primers suggests that homopolymeric stretches of dA's may have a negative impact on sequence quality, but that these stretches must be quite long, and near to the 3′ terminus to exert their effects. Other primers with extensive homolopolymeric stretches yielded good results (Table 3). Finally, when the length of the longest homopolymeric stretch in each of the primers was plotted against the number of errors in the sequence reaction (Figure 2E), no general trend was observed when reactions were performed with dRhodamine reagents. Together, these observations suggest that under optimal conditions using dRhodamine reagents, long homopolymeric stretches do not necessarily negatively

Table 3. Primers with Physico-Chemical Characteristics that Were Not Expected to Provide Good Sequencing Results

Primer#	Sequence			Errors[1]					
				dRhodamines				"old" Rhodamines	
				per 700 nt	total read	per 700 nt	total read	per 700 nt	total read
Homopolymeric stretches:		*length[2]:*							
86	TGACTCCCCTTCTGAAAA	4A/4C		5	6/754	4	9/803	--	58/685
87	CTCCCCTTCTGAAAAAAA	7A/4C		4	8/755	6	16/830	10	13/748
88	CCCTTCTGAAAAAAAAGA	8A/3C		FAILED		13	17/819	--	62/558
89	TTCTGAAAAAAAAGAACA	8A/2A		--	46/683	7	19/842	57	63/714
810	TGAAAAAAAAGAACACGT	8A/2A		3	9/786	3	11/790	11	13/711
876	ATTATATGAAAATCCCTC	4A/3C		2	7/797	nd	nd	nd	nd
877	ATATGAAAATCCCTCCCC	4C/4A		0	3/772	nd	nd	nd	nd
Melting temperature - %G+C:		*Tm:*	*%G+C*						
810Ab	CCGCGGGTGCGCCGCAAAGGGC	91°C	82	3	6/773	nd	nd	nd	nd
884	TCCCCGGGTGCGCCGCAA	83°C	78	1	3/724	6	17/766	10	22/766
880	TCCCTCCCCCATTCCGCG	75°C	72	5	11/771	5	31/867	15	15/703
839	TAAGGTAAAATACACACA	44°C	28	2	4/768	7	9/735	12	15/738
872	GTGTACACTACAATTATA	37°C	28	3	3/763	nd	nd	16	32/780
Secondary priming sites:		*score[3]:*							
882	CCCCATTCCGCGGGTGCG || ||||| ||||||| TGGTAAAGGC-CCCACGC	304/518		5	9/767	8	19/823	14	18/709
886	GGTCCCCCGCAAAGGGCC | |||| ||||||| TCGGTCGGC-CTTCCCGG	374/517		1	8/776	8	14/803	23	46/784
849b	CGCTTTGACTCCCCTTCTGA | | |||| ||||||| ||| GTGCAACTAAGGGGAAAACT	246/451		6	11/784	nd	nd	nd	nd
810Ab	CCGCGGGTGCGCCGCAAAGGGC | | |||| |||||| TTTGGTCGGTCGGC-CTTCCCG	190/643		3	6/773	nd	nd	nd	nd

[1]Errors and ambiguities per 700 nt read, or per total read length. For the dRhodamines, both the initial results and the results of the repeat analysis are shown.

[2]Length of the longest and second longest homopolymeric stretches in the primer.

[3]X/Y where X is the priming efficiency number generated by Oligo version 5.0, for potential secondary priming in the vector, and Y is the value for a perfect match with that primer. In general, the higher the score, the higher the stability and quality of the priming at that site. A score of >200 will often give false priming results according to the Oligo 5.0 manual.

nd, not determined

effect sequence results.

T_m and percent G+C also showed less impact than expected on the dRhodamine reactions (Figure 2, C and D and Table 3). Thus, primers with low T_m and percent G+C; e.g., 37°C/28% and 44°C/28% for primer Nos. 72 and 39, and others with high T_m/percent G+C; e.g., 75°C/72% and 83°C/78% for control primer Nos. 80 and 84 and 91°C/82% for submitted primer No. 4321b, each provided sequence at well under 1% error rate (Table 3). Secondary priming potential also exhibited less of an obstacle for dRhodamine sequencing than was anticipated. Primers with very high secondary priming scores at sites within the template provided excellent sequence data. Control primer Nos. 82 and 83, and submitted primer Nos. 10Ab and 49b, had high secondary priming scores, and in each case the homology encompassed the 3′ end of the primer (Table 3). Nonetheless, these primers yielded very high-quality results.

Finally, we compared the sequencing results generated with primers selected by software to the results of manually selected primers and to the arbitrarily selected control primers. There was no observable difference in the results obtained; i.e., the average

number of errors over a 700-nt sequencing window was 2.9, 3.1 and 3.3 for all submitted primers, for submitted primers selected using software and for the control primers, respectively. Moreover, no clear trends were observed when comparing the physico-chemical characteristics of the software-designed primers to the manually designed or the control primers (data not shown).

Old Rhodamines vs. dRhodamines

The old rhodamine dye reactions were more strongly impacted by primer physico-chemical characteristics than the dRhodamine dye reactions. As described above, the old rhodamine dye reactions were consistently poorer than the dRhodamine reactions (Table 2). Moreover, the former reactions seemed more sensitive to primer characteristics. Thus, reactions with higher error rates tended to be those with lower T_m, lower percent G+C and longer homopolymeric stretches (Figure 2, panels C–E). As with the dRhodamines, the old rhodamine reactions exhibited no discernable trend in relation to secondary priming potential (Figure 2F), or primer secondary structure (data not shown).

DISCUSSION

In this study, we examined the perceptions and strategies of participant core facility personnel for the design and selection of DNA sequencing and PCR primers. The results may reveal some unexpected misconceptions in the requirements for good sequencing primer design. Thus, most laboratories espouse primer design dogma suggesting that a sequencing primer should be 18–24 nt in length with a T_m = 52°–66°C, 40%–60% G+C, 1–2-bp 3′ GC clamps and little internal structure, homopolymeric sequence or potential for secondary priming. The percent G+C and secondary priming potential were rated as the most important characteristics, but T_m, GC clamps, primer secondary and tertiary structure, length and the presence of homopolymeric runs were also thought to be relevant. Moreover, previous studies have shown that secondary structure of the template in the primer target regions can effect priming efficiency (7,8).

With regard to primer design, we were surprised to find that the majority of investigators design most of their sequencing primers manually. Thus, in spite of the perceived importance of primer characteristics and the widespread availability of primer design software, approximately 55% of the participants in this study elected to design primers manually. We believe that this observation is the result of a combination of the facts that primer design software is not yet sufficiently "user-friendly" and that the primer physico-chemical characteristics are not as important as commonly believed. It would seem that the authors of primer design software could take note of this result and streamline and enhance their products so that they are more convenient, beneficial and practical to use. Clearly, however, there are many circumstances (e.g., templates with GC-rich regions or with unusual sequences or structures, genomic sequencing or the sequencing of very large templates with many repeat regions) in which it would be advisable to use appropriate primer design software to automate and expedite primer design and synthesis.

The results of the empirical sequencing analysis were surprising in that nearly all of the primers yielded data of extremely high quality. Excellent data was obtained in spite of widely varying primer T_m and percent G+C, the presence of nearly 50% of the primer as homopolymer and the presence of very strong secondary priming sites, characteristics considered by most respondents to be the critical for sequencing primers. This study confirms earlier observations (3) that there is a broad tolerance in these and other characteristics of sequencing primers. Other factors, such as template purity or quality or technical expertise, possibly play greater roles. In this study, the plasmid template was selected for absence of sequence extension obstacles and purified by double banding in CsCl-ethidium bromide isopycnic density gradients. Therefore, this template was extremely pure and optimal for sequencing. The primers were similarly highly purified. The reactions were performed in a single high-throughput sequencing facility under tightly

controlled conditions. Different results may be obtained using less carefully purified DNA templates with unusual sequences or structures or in less rigorously controlled sequencing operations.

Finally, the data generated using old rhodamine reagents was, in general, poorer in quality than that generated using the dRhodamine reagents. Moreover, the expected trends in comparing sequence performance and primer characteristics were more prominently observed using the old rhodamine reagents. We believe it possible, if not likely, that the greater uniformity in peak height generated by the newer sequencing reagents (19) is at least partially responsible for this difference. Thus, background peaks are less likely to interfere with base calling, lessening the overall effects of potential problems like secondary priming and stutter.

ACKNOWLEDGMENTS

We would like to thank the ABRF for sponsoring this study and its members for their participation. We also thank John Fritz of the VCU Health Sciences Computing Center for his assistance in electronic distribution of the study and anonymous receipt of the responses. The ABRF is a nonprofit organization devoted to promoting excellence in biotechnology core facilities. Information about the ABRF and its programs can be obtained from the ABRF Business Office (Tel: 301-571-8300) or the ABRF Web site **(http://www.abrf.org)**.

REFERENCES

1. **Adams, P.S., M.K. Dolejsi., G. Grills, S. Hardin, D.L. McMinimy, P. Morrison, J. Rush, S. Goff et al.** 1998. An analysis of techniques used to improve the accuracy of automated DNA sequencing of a GC-rich template: results from the 2nd ABRF DNA Sequence Research Group Study. J. Biomol. Techniq. *9*:9-18.
2. **Adams, P.S., M.K. Dolejsl, S. Hardin, S. Mische, B. Nanthakumar, H. Riethman, J. Rush and P. Morrisson.** 1996. DNA sequencing of a moderately difficult template: evaluation of the results from a *Thermus thermophilus* unknown test sample. BioTechniques *21*:678.
3. **Grayburn, W.S. and T.L. Sims.** 1998. Anchored oligo(dT) primers for automated dye terminator DNA sequencing. BioTechniques *25*:340-346.
4. **Hyndman, D., A. Cooper, S. Pruzinshky, D. Coad and M. Mitsuhashi.** 1996. Software to determine optimal oligonucleotide sequences based on hybridization simulation data. BioTechniques *20*:1090-1097.
5. **Ivanetich, K., L. Bibbs, R.L. Niece, N.D. Denslow, C.W. Naeve, M. Ruhde and L.H. Ericsson.** 1997. Commercial biotech instrumentation survey on performance, support, service. Genetic Engineering News *17*:17.
6. **Ivanetich, K., R.L. Niece, M. Rohde, E. Fowler and T.K. Hayes.** 1993. Biotechnology core facilities: trends and update. FASEB J. *7*:1109-1114.
7. **Kaczorowski, T. and W. Szybalski.** 1996. Cooperativity of hexamer ligation. Gene *179*:189-193.
8. **Kaczorowski, T. and W. Szybalski.** 1998. Genomic DNA sequencing by SPEL-6 primer walking using hexamer ligation. Gene *223*:83-91.
9. **Lathe, R.** 1986. Synthetic oligonucleotide probes deduced from amino acid sequence data: theoretical and practical considerations. J. Mol. Biol. *183*:1-12.
10. **Maniatis, T., E.F. Fritsch and J. Sambrook.** 1982. Molecular Cloning; A Laboratory Manual. CSH Laboratory Press, Cold Spring Harbor, NY.
11. **McGraw, R.A., E.K. Steffe and S.M. Baxter.** 1990. Sequence-dependent oligonucleotide-target duplex stabilities: rules from empirical studies with a set of twenty-mers. BioTechniques *8*:674-678.
12. **Naeve, C.W., G.A. Buck, R.L. Niece, R.T. Pon, M. Robertson and A.J. Smith.** 1995. Accuracy of automated DNA sequencing: a multi-laboratory comparison of sequencing results. BioTechniques *19*:448-453.
13. **PE Biosystems.** 1991. Extension primers: recommendations for sequence selection, synthesis, and purification. *In* User Bulletin 19. PE Biosystems, Foster City, CA.
14. **Pon, R.T., G.A. Buck, K.M. Hager, C.W. Naeve, R. L. Niece, M. Robertson and A.J. Smith.** 1996. Multi-facility survey of oligonucleotide synthesis and an examination of the performance of unpurified primers in automated DNA sequencing. BioTeclutiques *21*:680-685.
15. **Pon, R.T., G.A. Buck, R.L. Niece, M. Robertson, A.J. Smith and E. Spicer.** 1994. A survey of nucleic acid services in core laboratories. BioTechniques *17*:526-534.
16. **Ryehlik, W. and R.E. Rhoads.** 1989. A computer program for choosing optimal oligonucleotides for filter hybridization, sequencing and in vitro amplification of DNA. Nucleic Acids Res. *17*:8543-8551.
17. **Strauss, E.C., J.A. Kobori, G. Siu and L.E. Hood.** 1986. Specific-primer-directed DNA sequencing. Anal. Biochem. *154*:353-360.
18. **Vuss, H., S. Wiemann, D. Grothues, C. Sensen, J. Zimmermann, C. Schwager, J. Stegemann, H. Erfle, T. Rupp and W. Ansorge.** 1993. Automated low-redundancy large-scale DNA sequencing by primer walking. BioTechniques *15*:714-721.
19. **Zakeri, H., G. Amparo, S.-M. Chen, S. Spurgeon and P.-Y. Kwok.** 1998. Peak height pattern in dichloro-rhodamine and energy transfer dye terminator sequencing. BioTechniques *25*:406-414.

Received 20 January 1999; accepted 26 May 1999.

Address correspondence to:

Dr. Gregory A. Buck
Department of Microbiology and Immunology
P.O. Box 980678
Medical College of Virginia Campus
Virginia Commonwealth University
Richmond, VA 23298-0678, USA
Internet: buck@hsc.vcu.edu